The Successive Unconstrained Least-Squares Estimation Technique : SULSET (科学計算基本ライブラリーのアルゴリズム)

AUTHOR(S):

OKAMOTO, YOSHIHARU; ONG, SIAK-HOO

数理解析研究所講究録
115 巻 1971 年 94-117

THE SUCCESSIVE UNCONSTRAINED LEAST-SQUARES ESTIMATION TECHNIQUE (SULSET)

Yoshiharu Okamoto
and
Siak-Hoo Ong
The University of Tokyo

Abstract : An algorithm of the class of the SUMT to solve the nonlinear programming problems with inequality and/or equality constraints is described. We call this technique " the Successive Unconstrained Least-Squares Estimation Technique (SULSET) " . We propose the acceleration technique and derive the rate of convergence in this paper.

1. Introduction

In order to solve the nonlinear programming problem :

(A) minimize $f(x)$ subject to $g_i(x) \leqq 0$, $i=1,\dots,m$, and $x \in E^n$,

the SUMT (Sequential Unconstrained Minimization Technique) has been proposed. This method is based upon the property that , when the problem (A) is transformed into a sequence of unconstrained minimization problems of appropriate penalty functions , the convergent point of the sequence of the unconstrained minima is the optimal solution of the problem (A). Carrol first proposed the CRST (Created Response Surface Technique) in his well-known paper [4] and later Fiacco and McCormick established this technique theoretically and practically in [5] - [12]. They called it the SUMT.

Fiacco and McCormick first proposed the " interior point method " in [7]. This method can be stated by introducing the following penalty function :

$$(1) \quad P(x,r_k) = f(x) - r_k \sum_{i=1}^{m} 1/g_i(x) \quad ,$$

where $\{r_k\}$ is a monotonically decreasing scalar sequence which tends to zero. The sequence of the minima of (1) converges to the optimal solution x^* of the problem (A) as $r_k \to 0$ ($k \to \infty$).

On the other hand , Zangwill proposed the " exterior point method " in [28]. The penalty function can be defined as follows :

$$(2) \quad T(x,t_k) = f(x) + t_k \sum_{i=1}^{m} [\ \max(\ 0, g_i(x)\ )\ ]^{1+\varepsilon} \ , \varepsilon > 0 \quad .$$

The sequence of the minima of (2) converges to x^* as $t_k \to \infty$($k \to \infty$).

In the former method , the so-called " barriers " on the boundary of the feasible region , which the penalty functions construct , prevent the sequence of the minima $\{x^k\}$ from jumping out of the interior of the feasible region. While in the latter method , the so-called " walls " are constructed outside the feasible region and the sequence of the minima $\{x^k\}$ slips down to the optimal solution of the problem (A) along them.

Furthermore the generalizations of various types of penalty functions have been considered and the convergence properties have been also discussed by Fiacco and McCormick [11] , Fiacco [6] , Fiacco and Jones [12] and Stong [27].

Afterward Fiacco and McCormick proposed the SUMT without parameters in [9] and [11]. They called this method the " Q-function type SUMT ". The following penalty function was proposed :

$$(3) \quad Q(x,x^{k-1}) = 1/[\ f(x^{k-1}) - f(x)\ ] - \sum_{i=1}^{m} 1/g_i(x) \quad .$$

This method can be classified to an interior point type.

On the contrary , the Q-function type exterior point methods are proposed by Morrison [22] and Kowalik , Osborne and Ryan [18]. For the most part of this paper , we will discuss the method of this type and name this class of methods the " Successive Unconstrained Least-Squares Estimation Technique (SULSET) ".

Section 2 mainly gives the results of [22] and [18] , section 3 gives our acceleration technique , section 4 gives the comparison with the numerical results of [18] , section 5 gives the derivation of the rate of convergence , and section 6 gives the generalization of the SULSET.

For unconstrained minimizations , recently the conjugate gradient type methods are very prevalent ([14] , [3] , [13] , [15] and [16]). However for unconstrained least-squares estimations , we would like to recommend the Marquardt's Maximum Neighbourhood Method ([2] , [17] , [20] and [24]). Because this method converges rapidly and is very stable in characteristics .

2. The Successive Unconstrained Least-Squares Estimation Technique

Let us consider the following nonlinear programming problem :

(B) minimize $f(x)$ subject to $g_i(x)=0$, $i=1,\ldots,m$, and $x\in E^n$.

Morrison suggested the algorithm to transform the problem (B) to successive least-squares problems :

(C) minimize $S(x,X_k)$ subject to $x\in E^n$,

where

(1) $$S(x,X_k) = [\ f(x)-X_k\ ]^2+\sum_{i=1}^{m} c_i\ g_i(x)^2$$

c_i , $i=1,\ldots,m$, are positive constants and

$\{X_k\}$ is a monotonically increasing scalar sequence which tends to the minimum value v^* of the problem (B).

The next theorem shows the validity of the above transformation.

THEOREM 2.1 (Convergence Theorem of the SULSET) If (a) f and g_i , $i=1,\dots,m$, are continuous functions of x , (b) $S(x,X_k)$ is the function as defined in (1) , (c) $\{X_k\}$ is a monotonically increasing sequence and converges to v^* , (d) A^* is the set of the optimal solutions of the problem (B) , and (e) x^k is the minimum of $S(x,X_k)$, then

(i) $f(x^k) \leq v^*$,

(ii) $f(x^k) \leq f(x^{k+1})$,

(iii) there exists a limit of the sequence $\{x^k\}$ and this limit point belongs to A^* , and

(iv) $\lim_{k\to\infty} S(x^k,X_k) = 0$.

Proof. See the ref. [22] and the proof of Theorem 6.1 .

It is a question of how to construct the sequence $\{X_k\}$ to converge towards v^* from below. The Morrison-parameter sequence is following :

(2) $$X^M_{k+1} = X_k + [\ S(x^k,X_k)\ ]^{1/2} \quad .$$

In order to ensure that the Morrison-parameter sequence converges to v^* , the following assumption is necessary.

ASSUMPTION 2.1 (Continuity Condition) Let $v^*(q)$ be the minimum value of the problem :

minimize $f(x)$ subject to $g_i(x) = q_i$, $i=1,\dots,m$.

$v^*(q)$ is continuous at $q=0$.

This assumption is quite natural , because it means that if the constraints changes only a little bit , then the minimum value changes only a little bit. The following theorem shows that the Morrison-parameter is effective.

THEOREM 2.2 (Convergence Theorem of the Morrison-Parameter Sequence)

If the conditions of Theorem 2.1 and Assumption 2.1 are satisfied , the the Morrison-parameter sequence $\{X_k\}$, such as defined in (2) , converges to v* .

Proof. See the proof of Theorem 6.2 .

The relation between dual feasibility and the SULSET is stated in the following theorem .

THEOREM 2.3 Let the conditions of Theorem 2.1 be satisfied and $f(x)$ and $g_i(x)$, $i=1,\dots,m$, be continuously differentiable functions with respect to their arguments ; let (x^*,λ^*) be the stationary point of the Lagrangian

$$(3)\qquad L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i\, g_i(x)$$

corresponding to the solution x^* of the problem (B). If $g_i(x^k)$, $i=1,\dots,m$, are linearly independent for each k large enough , then

$$(4)\qquad c_i\, g_i(x^k)/[\ f(x^k)-X_k\ ] \rightarrow \lambda_i^* \quad \text{as } k\rightarrow\infty \quad \text{for } i=1,\dots,m .$$

Proof. See the proof of LEMMA 1 in [18] .

Furthermore Kowalik , Osborne and Ryan proposed the tangent parameter sequence :

$$(5)\qquad X_{k+1}^{T} = X_k + S(x^k,X_k)/\left|\ f(x^k)-X_k\right| \qquad .$$

The next theorem indicates the validity and effectiveness of tangent parameter sequence.

THEOREM 2.4 Let the conditions of Theorem 2.1 be satisfied. Then

(i) $X_{k+1}^{T} \geq X_{k+1}^{M}$

(ii) If the conditions of Theorem 2.3 hold and if there exists a set Φ containing (x^*, λ^*) in its interior such that

(a) L has the saddle point property in so that

$$L(x^*, \lambda) \leqq L(x^*, \lambda^*) \leqq L(x, \lambda^*) \quad , \text{ and}$$

(b) for $\bar{\lambda}$ such that $(x, \bar{\lambda}) \in \Phi$ there exists an $\bar{x}$ such that $(\bar{x}, \bar{\lambda}) \in \Phi$, $L(\bar{x}, \bar{\lambda}) \leqq L(x, \bar{\lambda})$ for all x such that $(x, \bar{\lambda}) \in \Phi$ and $(\bar{x}, \bar{\lambda})$ is the unique solution in Φ to the system of equations

$$\nabla L(x, \bar{\lambda}) = 0$$

then $v^* \geqq X_{k+1}^{T}$ provided X_k is sufficiaently close to v^*.

Proof. (i) It is obvious that $X_{k+1}^{T} \geqq X_k$. From the definition of x^k ,

$$(X_{k+1}^{T} - X_k) \left| f(x^k) - X_k \right| \geqq (f(x^k) - X_k)^2 \quad .$$

Therefore

(6) $$(X_{k+1}^{T} - X_k) \geqq \left| f(x^k) - X_k \right|$$

Since

$$(X_{k+1}^{T} - X_k) \left| f(x^k) - X_k \right| = S(x^k, X_k) \quad \text{and}$$
$$(X_{k+1}^{M} - X_k)^2 = S(x^k, X_k) \quad ,$$

from (6) ,

$$(X_{k+1}^{T} - X_k)^2 \geqq (X_{k+1}^{M} - X_k)^2 \quad .$$

Thus

$$X_{k+1}^{T} \geqq X_{k+1}^{M} \quad .$$

(ii) See the proof of Theorem 1 in [18] . Q.E.D.

For an inequality constraint $h(x) \leqq 0$, we will use the equivalent equality constraint

$$g(x) \doteqdot \max(0 , h(x)) = 0$$

By the above stated discussion , we can solve any constrained nonlinear programming problem.

3. The Accelerated Successive Unconstrained Least-Squares Estimation Technique

Morrison proposed the subfunctions such as (2.1) and gave X_{k+1} such as (2.2). This situation can be illustrated as Fig.3.1 .

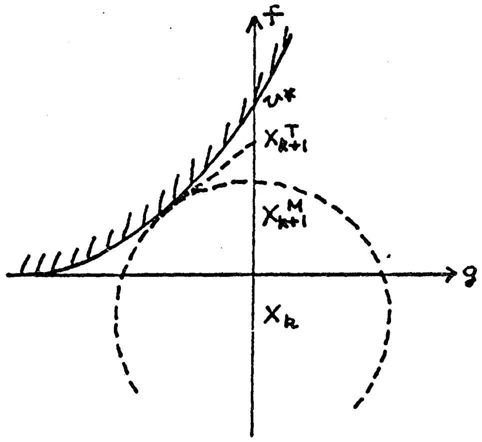

Fig.3.1

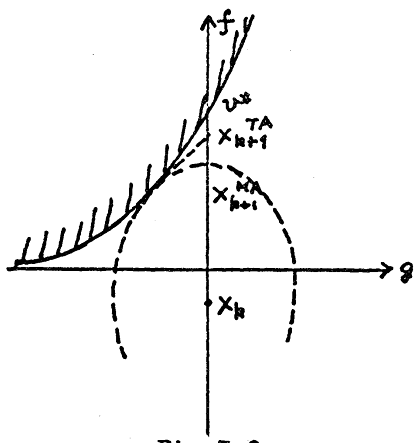

Fig.3.2

That is , the contours $(f-x)^2+g^2$=const. represent circles wirh the center (O,X) . If $S(x,X) = (f-X)^2 + t\, g^2$ ($t>1$) , the contours represent ellipses which are long in the derection f-axis such as Fig. 3.2 . Since the discussion in section 2 are all valid even if c_i , $i=1,\dots,m$, are different at each k , we will take the following system which is expected to have the more rapidly convergent property. We willpropose an algorithm to transform the problem (B) to the successive least-squares estimation problems :

(D) minimize $S(x,X_k,t_k)$

where

(1) $S(x,X_k,t_k) = [\ f(x)-X_k\ ]^2 + t_k \sum_{i=1}^{m} c_i\, g_i(x)^2$.

Let

(2) $X_{k+1}^{MA} = X_k + [\ S(x_A^k,X_k,t_k)\ ]^{1/2}$

(3) $t_{k+1} = c\, t_k$ ($c>1$, $t_k>0$) , and

(4) $X_{k+1}^{TA} = X_k + S(x_A^k, X_k, t_k)/\,|\,f(x_A^k) - X_k\,|$

where x_A^k is the minimum of the problem (D).

The following theorem shows that the above system has an accelerated convergence property.

THEOREM 3.1 If $t_k > 1$, then $X_{k+1}^{MA} \geq X_{k+1}^{M}$.

Proof. $X_{k+1}^{MA} - X_{k+1}^{M} = S(x^k, X_k)^{1/2} - S(x_A^k, X_k, t_k)^{1/2}$

$0 \leq S(x^k, X_k) \leq S(x_A^k, X_k) \leq S(x_A^k, X_k, t_k)$ Q.E.D.

We use the following algorithm which is like that of Kowalik et al..

ALGORITHM

0. Preparation Phase .

(i) Choose an arbitrary $x^s \in E^n$.

(ii) Move the solution to $x^o \in R$ by appropriate algorithms such that of [5] , where R is the feasible region.

I. Initial Phase .

(i) Set $X_o = f(x^o)$, k=1 and step = steplb .

(ii) Minimize $S(x, X_k, t_k)$ to find x^k . Here $X_k = X_{k-1}$ - step and $t_k=1$.

(iii) If $[\ S(x^k, X_k, t_k)\ ]^{1/2} <$ eps then step = 2 * step ; k=k+1 ; go to (ii) ; else BU= X_{k-1} ; proceed II .

II. Iteration Phase .

(i) Set BL = X_k and $t_{k+1} = c\ t_k$.

(ii) Compute X_{k+1}^{MA} , X_{k+1}^{TA} .

(iii) k=k+1 ; BL = X_k^{MA} . If $X_k^{TA} <$ BU , then $X_k = X_k^{TA}$ else $X_k = X_k^{MA}$.

(iv) Minimize $S(x, X_k, t_k)$ to find x^k .

(v) If $[\ S(x^k, X_k, t_k)\ ]^{1/2} \geq$ eps , then go to II(ii) else BU = X_k ; If BU-BL $<$ eps then go to end else $X_k \doteq$ BL and go to II(iv) .

It can be proved that this algorithm converges .

4. Numerical Results

In order to illustrate the acceleration performance stated in section 3 , let us show the numerical results of the following problems. In each table , (b) , (c) and (d) are the results showed in [18].

(i) Rosen-Suzuki problem :

$$f = x_1^2+x_2^2+2x_3^2+x_4^2-5x_1-5x_2-21x_3+7x_4$$

subject to

$$-x_1^2-x_2^2-x_3^2-x_4^2-x_1+x_2-x_3+x_4+8 \geq 0$$
$$-x_1^2-2x_2^2-x_3^2-2x_4^2+x_1+x_4+10 \geq 0$$
$$-2x_1^2-x_2^2-x_3^2-2x_1+x_2+x_4+5 \geq 0$$

The function has a minimum f=-44 at x=(0,1,2,-1).

(ii) Beale problem :

$$f = 9-8x_1-6x_2-4x_3+2x_1^2+2x_2^2+x_3^2+2x_1x_2+2x_1x_3$$

subject to $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$, $x_4 \leq 3$ and $x_4=x_1+x_2+2x_3$.

The function has a minimum f=1/9 at x=(4/3,7/9,4/9,3) .

(iii) Post office parcel problem :

$f = -x_1x_2x_3$ subject to $0 \leq x_1 \leq 20$, $0 \leq x_2 \leq 11$,

$0 \leq x_3 \leq 42$, $0 \leq x_4 \leq 72$ and $x_4 = x_1+2x_2+2x_3$.

The function has a minimum f=-3300 at x=(20,11,15,72) .

Table 4.1

THE ROSEN-SUZUKI PROBLEM

(a) The accelerated tangent parameter sequence (c=256)

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	-46.190235	0.09807343	1.0508771	2.1330896	-0.97308568
1	-44.392404	0.02013741	1.0103326	2.0257672	-0.98555903
2	-44.037648	0.00197659	1.0010217	2.0025327	-0.99835910
3	-44.002755	0.00014774	1.0000768	2.0001833	-0.99988065
4	-44.000182	0.00001172	1.0000049	2.0000108	-0.99999372
5	-44.000012	0.00000026	1.0000007	2.0000009	-0.99999925
6	-44.000000	0.00000007	1.0000000	2.0000000	-1.00000000

(b) The tangent parameter sequence

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	-44.810	0.05034	1.0236	2.0555	-0.9555
1	-44.029	0.00200	1.0010	2.0023	-0.9971
2	-44.000	0.00010	1.0000	1.9999	-1.0001
3	-44.010	0.00110	1.0005	2.0011	-0.9985
4	-44.009	0.00063	1.0003	2.0007	-0.9991
5	-44.005	0.00039	1.0002	2.0004	-0.9995
6	-44.003	0.00023	1.0001	2.0002	-0.9997
8	-44.001	0.00003	0.9999	2.0002	-0.9998
10	-44.000	0.00000	1.0000	2.0000	-0.9999

(c) The Morrison-parameter sequence

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	-44.810	0.050340	1.0236	2.0555	-0.9555
1	-44.460	0.031718	1.0147	2.0337	-0.9663
2	-44.273	0.019428	1.0090	2.0203	-0.9772
3	-44.160	0.011706	1.0054	2.0122	-0.9855
10	-44.004	0.000301	1.0001	2.0003	-0.9996
20	-44.000	0.000024	1.0002	1.9999	-1.0001

(d) The SUMT transformation

r	$f(x(r))$	x_1	x_2	x_3	x_4
10^0	-41.468	-0.01564	0.0125	1.8992	-0.8350
10^{-1}	-43.326	-0.00874	0.9709	1.9674	-0.9572
10^{-2}	-43.758	-0.00369	0.9933	1.9888	-0.9886
10^{-3}	-43.924	-0.00127	0.9985	1.9963	-0.9966
10^{-4}	-43.976	-0.00041	0.9996	1.9988	-0.9989
10^{-5}	-43.992	-0.00013	0.9998	1.9996	-0.9996
10^{-6}	-43.998	-0.00004	0.9995	1.9999	-0.9998
10^{-7}	-43.999	-0.00000	0.9999	2.0000	-0.9999

Table 4.2

THE BEALE PROBLEM

(a) The accelerated tangent parameter sequence (c=256)

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	0.08902419	1.2983708	0.80108687	0.50271733	3.1027244
1	0.11101519	1.3331901	0.77787364	0.44468401	3.0004219
2	0.11111105	1.3333333	0.77777774	0.44444461	3.0000003

(b) The tangent parameter sequence

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	0.0423	1.2054	0.8629	0.6572	3.3829
1	0.1104	1.3322	0.7785	0.4414	3.0034
2	0.1111	1.3336	0.7779	0.4442	3.0000
3	0.1111	1.3334	0.7778	0.4444	3.0000

(c) The Morrison-parameter sequence

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	0.0423	1.2054	0.8629	0.6572	3.3829
1	0.1091	1.3303	0.7798	0.4494	3.0089
2	0.1110	1.3333	0.7778	0.4446	3.0002
3	0.1111	1.3333	0.7779	0.4444	3.0001
4	0.1111	1.3333	0.7778	0.4445	3.0000

(d) The SUMT transformation

r	$f(x(r))$	x_1	x_2	x_3	x_4
10^0	0.7037	0.8952	0.7052	0.4285	2.4576
10^{-1}	0.2328	1.2929	0.6973	0.3314	2.6503
10^{-2}	0.1459	1.3793	0.7375	0.3516	2.8202
10^{-5}	0.1126	1.3356	0.7763	0.4407	2.9933
10^{-7}	0.1113	1.3336	0.7776	0.4441	2.9993
10^{-9}	0.1111	1.3333	0.7777	0.4444	2.9999

Table 4.3

THE POST OFFICE PARCEL PROBLEM

(a) The accelerated tangent parameter sequence (c = 64)

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	-3538.8889	20.381124	11.496137	15.103793	72.790492
1	-3537.1789	20.378465	11.492997	15.102620	72.784848
2	-3503.6666	20.325985	11.430291	15.080436	72.673719
3	-3317.4940	20.028554	11.041133	15.001921	72.057332
4	-3300.0180	20.000029	11.000043	15.000001	72.000059
5	-3300.0000	20.000000	11.000000	15.000000	72.000000

(b) The tangent parameter sequence

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	-3305.0	20.013	11.018	14.988	72.026
1	-3300.0	20.000	11.000	15.000	72.000

(c) The Morrison-parameter sequence

k	$f(x^k)$	x_1	x_2	x_3	x_4
0	-3305.0	20.013	11.018	14.988	72.026
5	-3304.8	20.012	11.018	14.988	72.025
30	-3304.1	20.010	11.015	14.990	72.021
100	-3302.5	20.006	11.009	14.994	72.013
200	-3301.8	20.004	11.007	14.996	72.010

(d) The SUMT transformation

r	$f(x(r))$	x_1	x_2	x_3	x_4
10^{0}	-3283.6	19.868	10.892	15.175	72.000
10^{-1}	-3294.8	19.959	10.964	15.057	72.000
10^{-2}	-3298.1	19.989	10.984	15.022	72.000
10^{-3}	-3299.0	19.997	10.989	15.012	72.000
10^{-5}	-3299.3	20.000	10.991	15.009	72.000
10^{-7}	-3300.0	20.000	11.000	15.000	72.000

From these numerical results , it can be said that our algorithm has the more rapidly convergent property than that of Kowalik et al. .

5. The Rate of Convergence

Let us derive the rates of convergence of the ordinary and accelerated Morrison-parameter sequences in the neighbourhood of the optimal solution x^* of the problem (B) .

If the sequence $\{a_n\}$ converges to α and if there exists real number ρ and $c \neq 0$ such that

$$\lim_{n\to\infty} \frac{| a_{n+1} - \alpha |}{| a_n - \alpha |^{\rho}} = c$$

then ρ is called the rate of convergence of $\{a_n\}$ and c is called the asymptotic error constant.

In order to prove the asymptotic convergence theorems of the SULSET , the next lemma is necessary.

<u>LEMMA 5.1</u> Let P be an n x k - matrix , the rank of which is m , H be a positive definite $n \times n$ - matrix and $A = P\,P^T$. Then there exists

(1) $\quad V = \lim_{t\to\infty} [\, tA + H \,]^{-1}$

and the rank of V is (n-l) , where $l = \min(\, n,m \,)$. And

(2) $\quad V\,P = 0$.

Furthermore , if $t \to \infty$, the approach of $[\, t\,A + H \,]^{-1}P$ towards null-matrix is elementwisely the order of $1/t$.

Proof. If $m \geq n$, the rank of A is n and A is positive definite. Therefore

(3) $\quad V = \lim_{t\to\infty} [\, t\,A + H \,]^{-1} = \lim_{t\to\infty} \frac{1}{t} [\, A + \frac{1}{t} H \,]^{-1} = \lim_{t\to\infty} \frac{1}{t} A^{-1} = 0$,

by which the conclutions are satisfied.

If $m < n$, let us form an $n \times m$ - matrix $\tilde{P}$ of rank m which consists of a set of m linearly independent column vectors of P and an $n \times (n-m)$ - matrix N which consists of a set of $(n-m)$ linearly independent column vectors which are orthogonal to all the column vectors of $\tilde{P}$. Then the $n \times n$ - matrix $[\ \tilde{P}\ N\ ]$ is nonsingular . Let

$$(4) \qquad \tilde{V}(t) = [\ \tilde{P}\ N\ ]^{-1}[\ t\ A + H\ ]^{-1}\begin{bmatrix} \tilde{P}^T \\ N^T \end{bmatrix}^{-1}$$

$$= \left[t \begin{bmatrix} \tilde{P}^T \\ N^T \end{bmatrix} [\ P\ P^T\ ][\ \tilde{P}\ N\ ] + \begin{bmatrix} \tilde{P}^T \\ N^T \end{bmatrix} H [\ \tilde{P} N] \right]^{-1}$$

$$= \left[t \begin{bmatrix} \tilde{P}^T P\ P^T \tilde{P} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \tilde{P}^T H\ P & \tilde{P}^T H\ N \\ N^T H\ \tilde{P} & N^T H\ N \end{bmatrix} \right]^{-1}$$

Let

$$(5) \qquad Q = \tilde{P}^T P\ P^T \tilde{P}\ , \quad Q_1 = \tilde{P}^T H\ \tilde{P}\ , \quad Q_2 = \tilde{P}^T H\ N\ , \quad Q_3 = \tilde{P}^T H\ N \text{ and } Q_4 = N^T H\ N\ .$$

Then by using the partial inversion of matrix ,

$$(6) \qquad \tilde{V}(t) = \begin{bmatrix} [\ t\ Q + Q_1 - Q_2 Q_4^{-1} Q_3\ ]^{-1} & -[\ t\ Q + Q_1\ ]^{-1} Q_2 [Q_4 - Q_3 [tQ+Q_1]^{-1} Q_2]^{-1} \\ -Q_4^{-1} Q_3 [tQ+Q_1 - Q_2 Q_4^{-1} Q_3]^{-1} & [Q_4 - Q_3 [tQ+Q_1]^{-1} Q_2]^{-1} \end{bmatrix}$$

Since $[\ tA + H\ ]$ is positive definite , $\tilde{V}(t)$ is positive definite. Therefore $[\ tQ+Q_1-Q_2Q_4^{-1}Q_3\ ]$ and $[\ Q_4-Q_3[\ tQ+Q_1\ ]^{-1}Q_2\ ]$ are positive definite and invertible. And since Q and Q_1 are positive definite , from the discussion of the case $m \geq n$,

$$(7) \qquad \lim_{t \to \infty} [\ tQ + Q_1 - Q_2 Q_4^{-1} Q_3\ ]^{-1} = 0 \quad , \text{ and}$$

$$(8) \qquad \lim_{t \to \infty} [\ tQ + Q_1\ ]^{-1} = 0 \quad .$$

Thus

$$(9) \qquad \tilde{V} = \lim_{t \to \infty} \tilde{V}(t) = \begin{bmatrix} 0 & 0 \\ 0 & Q_4^{-1} \end{bmatrix}$$

Since Q_4 is an $(n-m) \times (n-m)$ - positive definite matrix , the rank of $\tilde{V}$ is $(n-m)$. And since $[\ \tilde{P}\ N\ ]$ is nonsingular , the rank of V is $(n-m)$. From (4) ,

$$(10) \qquad V = [\ \tilde{P}\ N\ ]\ \tilde{V} \begin{bmatrix} \tilde{P}^T \\ N^T \end{bmatrix}$$
$$= N\ [\ N^T H\ N\ ]^{-1} N^T \quad .$$

Therefore , from the definition of N ,

$$V\ \tilde{P} = 0 \quad , \text{that is ,} \quad V\ P = 0 \quad .$$

Since

$$(11) \qquad V(t) = [\ tA + H\ ]^{-1} = [\ \tilde{P}\ N\ ]\ \tilde{V}(t) \begin{bmatrix} \tilde{P}^T \\ N^T \end{bmatrix} .$$
$$= \tilde{P}[\ tQ+Q_1-Q_2Q_4^{-1}Q_3\ ]^{-1}\tilde{P}^T - NQ_4^{-1}Q_3[\ tQ+Q_1-Q_2Q_4^{-1}Q_3\ ]^{-1}\tilde{P}^T$$
$$- \tilde{P}[\ tQ+Q_1\ ]^{-1}Q_2[\ Q_4-Q_3[tQ+Q_1]^{-1}Q_2\ ]^{-1}N^T +$$
$$N\ [\ Q_4 - Q_3[\ tQ+Q_1\ ]^{-1}Q_2\ ]^{-1}N^T \quad ,$$

$$(12) \qquad V(t)\ P = \tilde{P}\ [\ tQ + Q_1 - Q_2Q_4^{-1}Q_3\ ]^{-1}\tilde{P}^T\ P - NQ_4^{-1}Q_3[tQ+Q_1-Q_2Q_4^{-1}Q_3]^{-1}\tilde{P}^T P \ .$$

Therefore

$$\lim_{t\to\infty} t\ [\ tA + H\ ]^{-1}P = \tilde{P}Q^{-1}\tilde{P}^T P - NQ_4^{-1}Q_3Q^{-1}\tilde{P}^T P \quad ,$$

that is , if $t \to \infty$, the approach of $[\ tA + H\ ]^{-1}P$ towards null matrix is elementwisely the order of $1/t$. Q.E.D.

<u>THEOREM 5.1</u> (Asymptotic Convergence Theorem of the Morrison-parameter Sequence System) If the conditions of Theorem 2.1 , 2.2 , and 2.3 are satisfied and $\sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*)$ is positive definite or $P\ P^T$ is positive definite , then for k large enough , the rate of convergence of the Morrison-parameter sequence of the SULSET (C) ($c_i = 1$, $i=1,\ldots,m$) ,

$$(13) \qquad R_M = \frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|} \cong \frac{1}{2} K \quad , \text{ where}$$

$$(14) \qquad K = \lim_{t\to\infty} \nabla f(x^*)^T\ t[\ t[\ PP^T + \nabla f(x^*) \nabla f(x^*)^T] + \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*)]^{-1} \nabla f(x^*)$$

and

(15) $P = [\nabla g_1(x^*) \dots \nabla g_m(x^*)]$.

Proof. From the Taylor's theorem ,

(16) $f(x) = f(x^*) + \nabla f(x^* + \theta\delta)^T \delta$ and

(17) $g_i(x) = g_i(x^*) + 2\, g_i(x^*)\nabla g_i(x^*)^T \delta + \frac{1}{2}\delta^T[\, 2\nabla g_i(x^* + \theta_i\delta)$
$\nabla g_i(x^* + \theta_i\delta)^T + 2\, g_i(x^* + \theta_i\delta)\, \nabla^2 g_i(x^* + \theta_i\delta)\,]\,\delta$

where $x = x^* + \delta$, $0 \leq \theta, \theta_i \leq 1$ and $i=1,\dots,m$.

Since k is large enough , from Theorem 2.1 , 2.2 and 2.3 , x^k is near enough to x*. Therefore

$$\nabla f(x^* + \theta\delta) \cong \nabla f(x^*)$$

$$\nabla g_i(x^* + \theta_i\delta) \cong \nabla g_i(x^*) \text{ , } i=1,\dots,m \text{ ,}$$

$$\nabla^2 g_i(x^* + \theta_i\delta) \cong \nabla^2 g_i(x^*) \text{ , } i=1,\dots,m \text{ , and}$$

$$g_i(x^* + \theta_i\delta) \,/\, [\, f(x^*) - X_k\,] \cong \lambda_i^* \text{ , } i=1,\dots,m \text{ .}$$

From (2.1) , (15) and (17) ,

(18) $$\delta^k = -[\tau_k[\, PP^T + \nabla f(x^*)\nabla f(x^*)^T] + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)\,]^{-1}\nabla f(x^*)$$

where

(19) $$\tau_k = 1 \,/\, [\, f(x^*) - X_k\,] \text{ .}$$

Since τ_k is large enough , from Lemma 5.1 ,

(20) $$R_M \cong \frac{\|\delta^{k+1}\|}{\|\delta^k\|} \cong \frac{\tau_k}{\tau_{k+1}} = \frac{f(x^*) - X_{k+1}}{f(x^*) - X_k}$$
$$= 1 - [\, S(x^k, X_k)\,]^{1/2} \,/\, [\, f(x^*) - X_k\,]$$
$$= 1 - [\, 1 - \nabla f(x^*)^T \tau_k [\tau_k[\, PP^T + \nabla f(x^*)\nabla f(x^*)^T] + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)\,]^{-1}\nabla f(x^*)\,]^{1/2}$$
$$\cong \frac{1}{2} K$$

Q.E.D.

<u>THEOREM 5.2</u> (Asymptotic Convergence Theorem of the Accelerated Morrison-parameter Sequence System of the SULSET) If the conditions of Theorem

5.1 are satisfied , for k large enough , the asymptotic convergence property of the accelerated Morrison-parameter sequence system of the SULSET (D) can be described such as

$$(21)\qquad R_A = \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \cong \frac{M}{2c\ t_k}$$

where

$$(22)\qquad M = \lim_{t\to\infty} \nabla f(x^*)^T t\ [\ t\ PP^T + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)]^{-1} \nabla f(x^*)$$

Proof. By using the same formulation of the proof of Theorem 5.1 ,

$$(23)\qquad \delta^k = -[\tau_k[\ PP^T + \frac{1}{t_k}\nabla f(x^*)\nabla f(x^*)^T\ ] + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)]^{-1} \nabla f(x^*)$$

where

$$(24)\qquad \tau_k = \frac{t_k}{f(x^*)-X_k}$$

Therefore

$$(25)\qquad R_A \cong \frac{\|\delta^{k+1}\|}{\|\delta^k\|} = \frac{\tau_k}{\tau_{k+1}} = \frac{1}{c}\cdot\frac{f(x^*)-X_{k+1}}{f(x^*)-X_k}$$

where

$$(26)\qquad \frac{f(x^*)-X_{k+1}}{f(x^*)-X_k} = 1 - [\ 1 - \frac{1}{t_k}\nabla f(x^*)^T \tau_k[\tau_k[\ PP^T + \frac{1}{t_k}\nabla f(x^*)\nabla f(x^*)^T] + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)\ ]^{-1} \nabla f(x^*)\ ]^{1/2}$$

$$\cong \frac{M}{2\ t_k}$$

Q.E.D.

6. The Generalization

Let us give the genaralized formulation for the SULSET. For the problem (A) , we will take the following function :

$$(1)\qquad S(x,X_k) = |\ f(x)-X_k|^\alpha + W(x)\quad ,\quad \alpha > 0\ ,$$

where $W(x)$ is a continuous function with respect to x and

(2) $W(x) = \begin{cases} 0 & (\text{ if } g_i(x) \leq 0 \text{ , } i=1,\ldots m \text{ }) \\ >0 & (\text{ otherwise }) \end{cases}$

Let the feasible region be denoted by $R = \{ x \mid g_i(x) \leq 0 \text{ , } i=1, \ldots, m \}$ and

if $\inf_{x \in R} f(x) < X_k < \sup_{x \in R} f(x)$,

then $\inf_{x \in E^n} S(x,X_k) = 0$.

It can be proved that if there exists a solution of the problem (A) and the sequence $\{X_k\}$ converges to the minimum value v* from below , then under appropriate conditions , the sequence of minima , $\{x^k\}$, of $S(x,X_k)$ converges to the solution.

In order to prove the convergence property , the following definition and lemma are necessary.

DEFINITION 6.1 A nonempty set $M^* \subset M$ is called an isolated set of M if there exists a closed set E such that $E^0 \supset M^*$ and such that if $x \in E-M^*$ then $x \notin M$.

LEMMA 6.1 If a set of local minima A* corresponding to the local minimum value v* of the problem (A) is a nonempty isolated compact set , then there exists a compact set S such that $A^* \subset S^0$, and for any point $y \in R \cap S$, if $y \notin A^*$, then $f(y) > v^*$.

Proof. See the ref.[11] , p.47.

The proof of convergence follows the format used by Fiacco and McCormick.

THEOREM 6.1 (Convergence Theorem of the Generalized SULSET) If (a) f , g_1 ,..., g_m are continuous functions of x , (b) $S(x,X_k)$ is the function as defined in (1) , (c) a set of points A^* that are local minima of the problem (A) corresponding to the local minimum value v^* is a non-empty isolated compact set , and (d) $\{X_k\}$ is a monotonically increasing sequence which converges to v^* from below , then

(i) $f(x^k) \leq v^*$,

(ii) there exists a compact set S , given in Lemma 6.1 , such that $A^* \subset S^o$, and for k large enough , the unconstrained minima x^k of $S(x,X_k)$ in S^o exists and every limit point of any subsequence of $\{x^k\}$ is in A^* .

(iii) $\lim_{k\to\infty} W(x^k) = 0$;

(iv) $\lim_{k\to\infty} f(x^k) = v^*$;

(v) $\lim_{k\to\infty} S(x^k,X_k) = 0$.

Proof. Since

$$|f(x^k)-X_k|^\alpha \leq S(x^k,X_k) \leq S(x^*,X_k) = (\ v^*-X_k\ )^\alpha \quad , \text{ where } x^* \in A^* \quad ,$$

if $f(x^k) > X_k$ then $f(x^k)-X_k \leq v^*-X_k$, that is , $f(x^k) \leq v^*$ else $f(x^k) \leq X_k \leq v^*$. This proves part (i).

From the conditions (a) and (c) , there exists a compact set S where $A^* \subset S^o$ such that $f(y) > v^*$ for all $y \in R \cap S$ and $y \notin A^*$.

Let x^k be a minimum of $S(x,X_k)$ in S. Since $\{x^k\}$ is on the compact set S , there exists a convergent subsequence. For simplicity , let it aiso be denoted by $\{x^k\}$.

Let $\hat{x}$ be the convergent point of $\{x^k\}$ and let us assume $\hat{x} \notin A^*$. Then $S(x,v^*) > 0$. Since $\lim_{k\to\infty} X_k = v^*$, if $x^o \in A^*$, then $\lim_{k\to\infty} S(x^k,X_k) = S(\hat{x},v^*) > 0 = \lim_{k\to\infty} S(x^o,X_k)$. Therefore for k large enough , because of the continuity of $S(x,X_k)$ with respect to x and X_k , this contradicts the fact that x^k is the minimum of $S(x,X_k)$. Hence $\hat{x} \in A^*$. But since $A^* \subset S^o$, for k large enough , x^k must be in S^o. This proves part (ii). Parts (iii) - (v) follow from the fact that $\lim_{k\to\infty} f(x^k) = v^*$.

The Morrison's type parameter sequence is following in this case :

(3) $\quad X_{k+1}^{M} = X_k + [\ S(x^k, X_k)\ ]^{1/\alpha}$.

In order to prove the convergence of the Morrison-parameter sequence to v* from below , let us provide the following assumptions.

<u>ASSUMPTION 6.1</u> Let the minimum value of the problem :

minimize: $f(x)$ subject to $g_i(x) \leqq q_i,\ q_i \geqq 0$, $i=1,\dots,m$,

be a function of m-vector $q = (\ q_1,\dots,q_m\ )^T$ and be denoted by $v^*(q)$. Then $v^*(q)$ is continuous at $q = 0$ in the region $\{\ q \mid q_i \geqq 0, i=1,\dots,m\}$.

<u>ASSUMPTION 6.2</u> For $\forall \varepsilon > 0$ there exusts $\exists \Delta(\varepsilon) > 0$ such that

$$\sum_{i=1}^{m} [\ \max(\ 0,\ g_i(x)\ )\ ]^2 < \varepsilon \quad \text{for } \forall x \in \{x \mid W(x) < \Delta(\varepsilon)\} \ .$$

<u>THEOREM 6.2</u> (Convergence Theorem of the Generalized Morrison-parameter Sequence) If the conditions of Theorem 6.1 , Assumption 6.1 and 6.2 are satisfied , then the Morrison-parameter sequence $\{X_k\}$, such as defined in (3) , converges to v* (= v*(0)) from below.

Proof. If $X_k \leqq v^*$, then $X_{k+1} \leqq v^*$, because $X_{k+1} \geqq X_k$ and

$$\begin{aligned}(\ X_{k+1}-X_k\ )^{\alpha} &= S(x^k, X_k) \\ &\leqq |f(x^*)-X_k|^{\alpha} + W(x^*) \quad , \ \forall x^* \in A^* \ , \\ &= |\ v^* - X_k\ |^{\alpha}\end{aligned}$$

Since $\{X_k\}$ is monotonically increasing and bounded above , it converges to X . From Assumption 6.1 , for $\forall \varepsilon > 0$ there exists $\delta > 0$ such that $|v^*(q)-v^*(0)| < \varepsilon$ for $\forall q = \{\ q \mid \|q\| < \delta\ ,\ q_i \geqq 0\ , i=1,\dots,m\}$. Let k be large enough so that $X_{k+1}-X_k < [\ \Delta(\delta)\ ]^{1/\alpha}$. Then $S(x^k, X_k) < \Delta(\delta)$. Therefore $W(x^k) < \Delta(\delta)$. From Assumption 6.2 , $\|q^k\| < \delta$ where $q_i^k = \max(\ 0, g_i(x)\ )$, $i=1,\dots,m$. So $|v^*(q^k)-v^*(0)| < \varepsilon$. Since $v^*(q^k) \leqq v^*(0)$, $v^*(q^k) > v^*(0)-\varepsilon$.

From the definition of $v^*(q^k)$, $v^*(q^k) \leqq f(x^k)$.

Hence $0 \leq v^*(0) - f(x^k) < \varepsilon$.

Therefore $f(x^k) \to v^*$. Thus $X_k \to v^*$ as $k \to \infty$, i.e. $X = v^*$.

Q.E.D.

For the accelerated SULSET of section 3 , a generalized expression can be described as follows :

$$(4) \qquad S(x, X_k, t_k) = | f(x) - X_k |^{\alpha} + p(t_k)\, W(x) \quad , \alpha > 0$$

where $p(t)$ has the property that if $0 < t_1 < t_2$, then $0 < p(t_1) < p(t_2)$ and if $\{ t_k \}$ is a monotonically increasing sequence of positive values where $\lim_{k \to \infty} t_k = \infty$, then $\lim_{k \to \infty} p(t_k) = \infty$.

Here it can be said that if $X_k \to v^*$ and $t_k \to \infty$, the sequence of minima of $S(x, X_k, t_k)$ converges into A^* and its convergence is accelerated.

7. Conclusions

Comparing with the Fiacco and McCormick's algorithm , this type of constrained minimization technique is better. The rate of convergence depends mainly upon the strategies of how the parameter X_k being chosen and the types of the penalty functions being used. The suggested acceleration technique performs excellently in practice. Our SULSET routine can solve any nonlinear programming problem with an arbitrary starting point. It automatically adjusts the solution to jump into the feasible region. From that point , in the case that the optimal solution is on the boundary of the feasible region , it moves the solution towards the boundary , where the optimal solution is located , until the solution begins to function until the grobal optimal solution is obtained. In the case that the optimal solution is inside the feasible region , after the value of X_k is less than the optimal solution value , as in the former case , the parameter sequence routine functions to give the optimal solution in a single iteration.

Several techniques are under consideration to improve the convergence and numerical stability property of the SULSET routine. An extension of applying the convergence theorem of Lagrange multiplier to our technique is also under consideration.

And our algorithm uses the Maximum Neighbourhood Method which is believed to be the best least-squares estimator. Therefore the computation is very stable.

At last , we would like to point out that the SULSET will be the leading technique for constrained minimizaton problems.

Acknowlegements

We would like to thank Professor Sigeiti Moriguti , Assistant Professor Masao Iri , Mr. T. Takasawa and members of Moriguti Laboratory of the University of Tokyo for their valuable suggestions and helps.

REFERENCES

[1] Allran,R.R. and Johnson,S.E.,"An algorithm for solving nonlinear programming problems subject to nonlinear inequality constraints" ,Comp.J.,Vol.13,No.2(1970),pp.171-177.

[2] Bard,Y.,"Comparison of Gradient Methods for the Solution of Nonlinear Parameter Estimation Problems",SIAM J. Numer.Anal.,Vol.7,No.1 (1970),pp.157-186.

[3] Broyden,C.G.,"Quasi-Newton Methods and their Application to Function Minimization",Math.Comp.,Vol.21(1967),pp.368-381.

[4] Carrol,C.W.,"The Created Response Surface Techniquefor Optimizing Nonlinear,Restrained Systems",Oper.Res.,Vol.9.No.2(1961),pp.169-184.

[5] Fiacco,A.V.,"Comments on the paper of C.W.Carrol",Oper.Res.,Vol.9, No.2(1961),pp.184-185.

[6] -----------,"A General Regularized Sequential Unconstrained Minimization Technique",SIAM J. Appl.Math.,Vol.17.No.6(1969),pp.1239-1245.

[7] Fiacco,A.V. and McCormick,G.P.,"The Sequential Unconstrained Minimization Technique for Nonlinear Programming , A Primal-Dual Method ",Man.Sci.,Vol.10,No.2(1964),pp.360-366.

[8] ------------------------------,"Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming",Man.Sci.,Vol.10,No.4(1964),pp.601-617.

[9] ------------------------------,"The Sequential Unconstrained Minimization Technique (SUMT) without Parameters",Oper.Res.,Vol.15,No.5(1967),pp.820-827.

[10] ------------------------------,"Extensions of SUMT for Nonlinear Programming : Equality Constraints and Extrapolation",Man.Sci.,Vol. 12,No.11(1966),pp.816-827.

[11] ------------------------------,"Nonlinear Programming:Sequential Unconstrained Minimization Technique",JohnWilley and Sons,Inc.(1968)

[12] Fiacco,A.V. and Jones,A.P.,"Generalized Penalty Methods in Topological Spaces",SIAM J.Appl.Math.,Vol.17,No.5(1969),pp.996-1000.

[13] Fletcher,R.(ed.),"Optimization",Academic Press(1969).

[14] Fletcher,R. and Powell,M.J.D.,"A rapidly convergent descent method for minimization",Comp.J.,Vol.6(1963),pp.33-41.

[15] Goldfarb,D.,"Extension of Davidon's Variable Metric Method to Minimization under Linear Inequality and Equality Constraints",SIAM J.

Appl.Math.,Vol.17,No.4(1969),pp.739-764.

[16] Goldfarb,D. and Lapidus,L.,"Conjugate Gradient Method for Nonlinear Programming Problems with Linear Constraints",I&EC Fundamentals,Vol. 7(1968),pp.142-151.

[17] Hartley,H.O.,"The Modified Gauss-Newton Method for the Fitting of Non-linear Regression Functions by Least Squares",Technometrica,Vol. 3,No.2(1961),pp.269-280.

[18] Kowalik,J.,Osborne,M.R. and Ryan,D.M.,"A new method for constrained optimization problems",Oper.Res.,Vol.17,No.6(1969),pp.973-983.

[19] Kuhn,H.W. and Tucker,A.W.,"Non-linear Programming",Proceeding of the Second Symposium on Mathmatical Statistics and Probability, J. Neyman(ed.),University of California Press,Berkeley(1951),pp.481-493.

[20] Marquardt,D.W.,"An algorithm for least-squares estimation of non-linear parameters",SIAM J.Appl.Math.,Vol.11,No.2(1963),pp.431-441.

[21] 森口繁一, 伊理正夫他"各種輸送網の将来計画の基本原理とその手法", 日科技連・計算センター (1970).

[22] Morrison,D.D.,"Optimization by least squares",SIAM J.Numer.Anal., Vol.5,No.1(1968),pp.83-88.

[23] Okamoto,Y.,Research Notes,Moriguti Laboratory,Department of Math. Engin.,The Univ. of Tokyo(September-October,1970).

[24] Ong,S.H.,"Techniques for nonlinear parameter estimation",Graduate Thesis,Department of Math. Engin.,The Univ. of Tokyo(1970).

[25] Powell,M.J.D.,"A survey of numerical methods for unconstrained optimization",SIAM J. Review,Vol.12,No.1(1970),pp.79-97.

[26] Spang,H.A.III.,"A review of minimization technique for nonlinear functions",SIAM J. Review,Vol.4(1962),pp.343-365.

[27] Stong,R.E.,"A note on the sequential unconstrained minimization technique for non-linear programming",Man.Sci.,Vol.12,No.1(1965), pp.142-144.

[28] Zangwill,W.I.,"Non-linear Programming via Penalty Functions",Man. Sci.,Vol.13,No.5(1967),pp.344-358.

[29] -------------,"Convergence Conditions for Nonlinear Programming Algorithm",Man.Sci.,Vol.16,No.1(1969),pp.1-13.

[30] -------------,"Nonlinear Programming:A Unified Approach",Prentice-Hall,Inc.(1969).